

ANALISIS SENTIMENT KOMENTAR VIDEO YOUTUBE “EPIC RAP BATTLE OF PRESIDENCY 2024” MENGGUNAKAN ALGORITMA NAÏVE BAYES & SVM

Rafif Dhia Yusrana¹, Gavin Berylian Josepto², Marvin Donald Richardo Aronggear³, Viktor Handrianus Pranatawijaya⁴, Ressa Priskila⁵.

^{1,2,3,4,5}, Teknik Informatika, Universitas Palangka Raya

email: ¹rafif0856@gmail.com, ²gavinberlylian18@gmail.com, ³marvinaronggear01@gmail.com,

⁴viktorhp@it.upr.ac.id, ⁵ressa@it.upr.ac.id

Abstract

Sentiment analysis of YouTube video recordings has become important in this increasingly technologically advanced era to understand users' reactions and opinions on various topics, including politics. Rapfight videos, especially those related to the presidential election, are one type of content that often attracts attention. This research uses the Naïve Bayes algorithm to assess the sentiment of comments on the video "Epic Rap Battle of Presidency 2024" on the YouTube site. The algorithms used in this research are Naïve Bayes Classifier and Support Vector Machine (SVM). There are five processes in this research, namely collecting YouTube comment data, preprocessing, labeling, classification, implementation and testing. From 3650 comment data on YouTube regarding the video "Epic Rap Battle of Presidency 2024" based on the results of the analysis, it was found that 31% of the comments were positive, 7% of the comments were negative and 62% of the comments were neutral. The accuracy results using the Naïve Bayes Classifier algorithm were 84.82% and the accuracy results using the Support Vector Machine (SVM) algorithm got the best results at 93.3%.

Keywords: You Tube Maximum, rapfight, Epic Rap Battle of Presidency 2024, Naïve Bayes, Support Vector Machine(SVM)

1. PENDAHULUAN

Zaman Sekarang ditandai dengan kemajuan teknologi informasi yang dimana memungkinkan konektivitas yang luas di seluruh dunia. Segala sesuatu didunia dapat terhubung dengan mudah memanfaatkan teknologi informasi yang berkembang pesat, sehingga kehidupan sehari-hari terasa lebih cepat, jarak terasa lebih pendek, segala hal terasa lebih mudah[1]. Bukan hanya itu saja, Terdapat dampak positif dari adanya kemajuan teknologi informasi ini, Dampak positif dari teknologi

informasi adalah contoh dapat melakukan komunikasi jarak jauh, dapat melakukan pencarian informasi dengan mudah[2]. Namun dampak negatifnya juga tidak bisa diabaikan. Salah satu dampak negatif yang terjadi adalah munculnya hate speech atau ujaran kebencian[3]. Hate speech adalah suatu tindakan yang dilakukan seseorang atau lebih dalam bentuk hasutan, provokasi maupun hinaan kepada orang lain dalam beberapa aspek seperti suku, agama, ras, gender, warna kulit, cacat, orientasi seksual, dan masih banyak lagi[4]. Di era digital ini, hate speech

semakin mudah tersebar dan dilontarkan melalui berbagai media, termasuk situs penyedia layanan streaming video seperti YouTube. YouTube adalah sebuah situs website media sharing video online terbesar dan paling populer di dunia internet. Saat ini pengguna youtube tersebar di seluruh dunia dari berbagai kalangan usia, dari tingkat anak-anak sampai dewasa[5]. YouTube menyediakan fitur komentar agar dapat dilakukannya komunikasi dua arah atau lebih oleh sesama pengguna. Namun, fitur kolom komentar juga seringkali digunakan sebagai tempat terjadi perilaku hate speech atau cyberbullying[6]. Pengguna seringkali menggunakan kolom komentar untuk menyampaikan pendapat yang berlebihan, merendahkan, atau bahkan merugikan orang lain.

Penelitian sebelumnya telah mengungkap sejumlah masalah dalam analisis sentimen komentar YouTube, seperti keterbatasan kamus sentimen, kesulitan dalam memberi label pada komentar dengan benar, dan kesulitan dalam mencapai kinerja klasifikasi yang memadai. Meskipun demikian, sejumlah penelitian menunjukkan bahwa menggabungkan kamus sentimen dan teknik algoritma klasifikasi dapat meningkatkan efektivitas deteksi sentimen pengguna dari komentar YouTube. Algoritma Naive Bayes merupakan salah satu yang dapat digunakan untuk klasifikasi teks dengan akurasi yang lumayan tinggi yaitu sebesar 89%[7]. Lalu algoritma support vector machine(SVM) memiliki nilai AUC yang lebih baik daripada naive bayes[8]. Penelitian ini mencoba menerapkan teknik ini dalam mengidentifikasi sentimen komentar pada video YouTube, khususnya pada konten "Epic Rap Battle of Presidency 2024", dengan tetap mempertimbangkan potensi dan manfaat algoritma Naive Bayes dan algoritma support vector machine(SVM).

Lalu Membandingkan kemudian mencari tahu apakah sebagian besar komentarnya positif atau negatif dan menilai keakuratan klasifikasi adalah tujuan utamanya. Oleh karena itu, penelitian ini merupakan langkah awal untuk memahami opini publik terhadap konten politik yang disajikan secara imajinatif, seperti rapfight di YouTube.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah proses memahami dan mengolah data tekstual secara otomatis untuk

mendapatkan informasi. Analisis sentimen dilakukan untuk mendeteksi opini terhadap suatu subjek dan objek (misalnya individu, organisasi ataupun produk) dalam sebuah kumpulan data[9].

2.2. Youtube

YouTube adalah platform media sosial yang memuat online video atau didalam YouTube terdapat berbagai macam video yang disimpan di dalam penyimpanan online yang berupa platform YouTube, sehingga video tersebut dapat dilihat dan diakses oleh siapapun dan dimanapun asalkan terdapat akses ke internet[10].

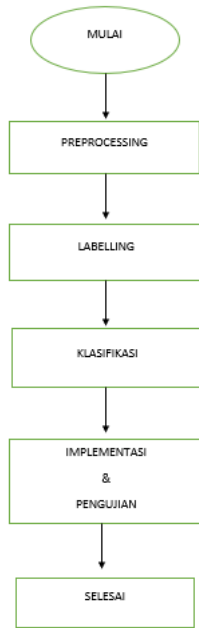
2.3. Algoritma Naïve Bayes dan Support Vector Machine

Algoritma naïve bayes adalah algoritma pembelajaran mesin untuk masalah klasifikasi yang terutama digunakan untuk klasifikasi teks yang melibatkan kumpulan data pelatihan berdimensi tinggi, Secara singkat algoritma naïve bayes classification adalah pengklasifikasi kumpulan data statistik yang mana untuk memprediksi semua probabilitas tiap anggota suatu class[11]. kemudian Support Vector Machine merupakan sistem pembelajaran yang menggunakan hipotesis berupa fungsi-fungsi linear dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan dengan teori optimasi[12].

Beberapa penelitian telah dilakukan untuk melakukan analisa dengan algoritma Naive Bayes dan Support Vector Machine seperti penelitian dari[13]. yang mendapatkan hasil akurasi dengan menggunakan algoritma naive bayes classifier sebesar 78,96%, dan menggunakan algoritma Support Vector Machine yang mendapatkan hasil akurasi sebesar 75,29%. Penelitian lainnya yaitu dari[14]. yang mendapatkan hasil akurasi dengan menggunakan algoritma Naive Bayes classifier sebesar 78,65%, dan menggunakan algoritma Support Vector Machine mendapatkan hasil akurasi sebesar 91,66%.

3. METODE PENELITIAN

Bagian Metode penelitian disini disebut sebagai tahapan pengerjaan yang dilakukan peneliti supaya tersusun dengan baik[15]. Berikut gambar tahapannya;



Gambar 1. Tahapan Metode

3.1. Pengumpulan Data Komentar Youtube

Tahap ini diperlukan pengumpulan data agar menganalisa dan memperoleh data serta informasi yang berhubungan terhadap penelitian[15]. Tahap pertama yaitu crawling yaitu mengumpulkan 3650 Data komentar pada video "Epic Rap Battle of Presidency 2024" di platform YouTube dikumpulkan menggunakan YouTube Data API atau teknik web scraping.

3.2. Preprocessing

Preprocessing pada data untuk membersihkan dan mempersiapkannya sebelum perhitungan, Setelah data komentar terkumpul, dilakukan tahap preprocessing untuk membersihkan data dari noise dan informasi yang tidak relevan. Preprocessing meliputi langkah-langkah seperti cleaning, tokenisasi, stopwords, normalisasi, dan stemming. Berikut contoh tahap preprocessing data dapat di lihat pada gambar dibawah:

	texts	english_text	cleaned_text	tokens	filtered_tokens	stemmed_tokens
0	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?
1	Terima kasih B2020 sudah memberikan informasi ini.	Thank you B2020 for providing this info.	thank you can for providing this info	thank you can for providing this info	thank you can for providing this info	thank you can for providing this info
2	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?
3	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?
4	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?

Gambar 2. Tahapan Preprocessing

3.3. Labelling

Tahap selanjutnya yaitu labelling, menggunakan NLP Vader_lexicon yang

bertujuan memberikan skor awal yang nantinya akan menentukan positif atau negatifnya suatu komentar. Proses labelling adalah proses memberikan label pada setiap kalimat. Tahap labelling dilakukan dengan melakukan perhitungan frekuensi kata yang muncul untuk diberi label positif atau negatif[16]. Melakukan pelabelan kalimat atau emoticon positif, dan negatif sebagai dataset latih. Setelah mengetahui pola kata positif dan negatif, kemudian menerapkan data set latih ke dataset uji dari live record di komentar youtube[1]. Berikut contoh tahap Labelling komentar dapat dilihat pada gambar dibawah.

english_text	cleaned_text	tokens	filtered_tokens	stemmed_tokens	sentiment_score	sentiment_positive	sentiment_negative	sentiment_neutral	sentiment_compound	sentiment_label
Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	Siapa yang akan memenangkan pemilu 2024? Apakah Biden atau Trump?	0.313	0.0	0.007	0.023	0.303	positive
Thank you B2020 for providing this info.	Thank you B2020 for providing this info.	Thank you B2020 for providing this info.	Thank you B2020 for providing this info.	Thank you B2020 for providing this info.	0.744	0.0	0.256	0.760	0.744	positive

Gambar 3. Tahapan Labelling

Setelah tahap labelling, dilanjutkan dengan tahap balanced distribution yang bertujuan untuk membagi data dengan klasifikasi positif, negative, dan netral. Data awal dibagi menjadi dua set, yaitu (X_train, y_train) yang merupakan data latih, kemudian (X_test, y_test) yang merupakan data uji.

3.4. Klasifikasi

Pada tahap klasifikasi ini dilanjutkan dengan menggunakan metode algoritma naïve bayes dan Support vector machine(SVM). Data yang sebelumnya telah melewati beberapa proses yaitu proses crawling dan labelling selanjutnya akan diproses menggunakan naïve bayes dan SVM. Naïve bayes dikembangkan oleh Reverend Thomas Bayes pada abad 18-an. Naïve bayes juga dikenal sebagai pengklasifikasian yang dapat diandalkan dan digunakan untuk memprediksi suatu class[17]. Diperkenalkan oleh Vapnik pada tahun 1992 bersama rekannya Bernhard Boser dan Isabelle Guyon. SVM merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam hal ini dimensi baru, akan mencari hyperplane untuk memisahkan secara linier dan dengan pemetaan nonlinier yang tepat ke dimensi

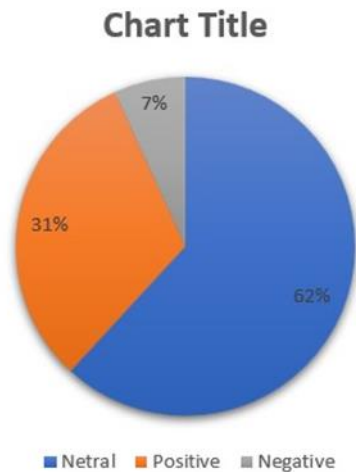
lebih tinggi, data dari dua kelas selalu dapat dipisahkan dengan hyperplane tersebut[18].

3.5. Implementasi & Pengujian

Selanjutnya pada tahap ini dilakukan dengan menggunakan Bahasa pemrograman python yang akan dijalankan pada platform google collabs. Tahap ini bertujuan untuk proses dilakukannya crawling, preprocessing, TF-IDF, yang merupakan identifikasi dari metode naïve bayes dan metode SVM yang telah berhasil dibentuk menggunakan Confusion matrix yang menghitung accuracy, precision, dan F1.

4. HASIL DAN PEMBAHASAN

Pada tahap processing data, telah dilakukan analisis terhadap komentar yang ditulis oleh penonton pada video youtube "Epic Rap Battle of Presidency 2024". Komentar tersebut telah dikelompokkan menjadi 3 jenis komentar yang telah di labelling yaitu komentar positif, negative, dan netral. Dari 3650 komentar ditemukan bahwa 31% komentar positif, 7% komentar negative, 62% komentar netral, bisa dilihat pada gambar dibawah ini.



Gambar 4. Data komentar yang telah dilabelling

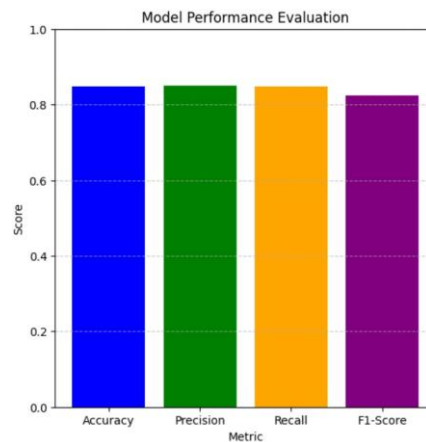
Pada analisis sentiment ini dilakukan 2 metode. Metode yang pertama yaitu menggunakan metode naïve bayes, Metode kedua yaitu menggunakan metode Support Vector Machine(SVM). Berikut tabelnya;

Naïve Bayes	
Accuracy	84,82%
Precision	85,07%
Recall	84,82%
F1-Score	82,45%

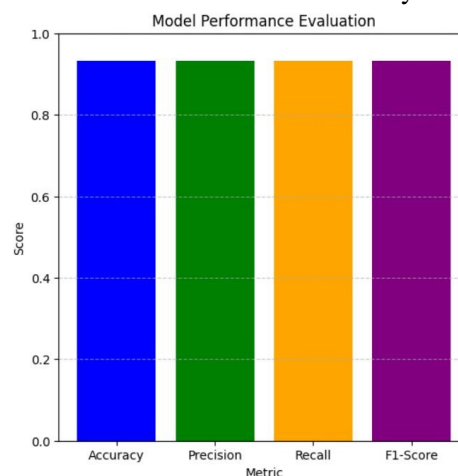
Tabel 2. Tabel Metode Support Vector Machine

Support Vector Machine	
Accuracy	93,3%
Precision	93,17%
Recall	93,3%
F1-Score	93,18%

Berdasarkan pada dua metode yang telah dilakukan sebelumnya, kedua metode tersebut menghasilkan accuracy yang paling besar yaitu 93,3% yang mana yang paling besar adalah accuracy metode Support Vector Machine(SVM) dibandingkan dengan metode Naïve Bayes. Berikut gambarnya



Gambar 5. Metode Naive Bayes



Gambar 6. Metode SVM

5. KESIMPULAN

Kesimpulan Dalam Analisis sentiment komentar video "EPIC RAP BATTLE OF PRESIDENCY 2024" di Youtube menunjukkan bahwa mayoritas komentar memiliki sentiment netral (62%), diikuti dengan positive (31%) dan negatif (7%). Algoritma SVM terbukti lebih efektif dalam klasifikasi sentiment dibandingkan Naïve

Bayes, dengan tingkat akurasi 93,3% dibandingkan 84,82%. Penelitian ini menunjukkan bahwa sentimen netral mendominasi persepsi publik terhadap video tersebut. Algoritma SVM dapat menjadi pilihan yang lebih baik untuk analisis sentimen dalam konteks ini..

6. REFERENSI

- [1] A. Rahman, F. Rahmat, M. Y. Fariqi, and S. Adi, "Metode Naive Bayes untuk Menganalisis Akurasi Sentimen Komentar di Youtube," 2020. [Online]. Available: <http://bit.ly/2u802Pe>
- [2] B. Jie, E. Eric, D. Mervyn, V. Anggrianto, K. Kelvin, and C. Gabriella, "Pemanfaatan Dan Dampak Penggunaan Teknologi Informasi Pada Bidang Sosial," *J. Inf. Syst. Technol.*, vol. 4, no. 2, pp. 392–397, 2023.
- [3] N. Rahmad, D. Setiyawan, D. Novita, R. D. Prasojo, and P. Hukum, "Jurnal SUTASOMA | 37 Tingkat Pemahaman Masyarakat Terhadap Tindak Pidana (Hate Speech, Hoax) Dalam Undang-Undang Informasi Dan Transaksi Elektronik Didesa Pekuncen Kec. Sempor." [Online]. Available: <https://s.id/jurnalsutasoma>
- [4] C. Oktiawan, "YURIDIS TINDAK PIDANA UJARAN KEBENCIAN DALAM MEDIA SOSIAL."
- [5] M. Arham, "Efektivitas penggunaan youtube sebagai media pembelajaran," *Acad. Educ.*, vol. 1, no. 1, pp. 1–13, 2020.
- [6] A. A. Ratnadewati, "CYBERBULLYING SEBAGAI PELANGGARAN ETIKA KOMUNIKASI DI MEDIA SOSIAL (ANALISIS KUANTITATIF PADA KOLOM KOMENTAR CHANNEL YOUTUBE RICIS OFFICIAL)," 2022.
- [7] M. Alhajji, A. Al Khalifah, M. Aljubran, and M. Alkhalifah, "Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19," 2020.
- [8] K. Fithriasari, R. W. Mayasari, N. Iriawan, and W. S. Winahju, "Surabaya Government Performance Evaluation using Tweet Analysis," *MATEMATIKA*, pp. 31–42, 2020.
- [9] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020.
- [10] R. P. Tutiasri, N. K. Laminto, and K. Nazri, "Pemanfaatan youtube sebagai media pembelajaran bagi mahasiswa di tengah pandemi covid-19," *J. Komunikasi, Masy. dan Keamanan*, vol. 2, no. 2, 2020.
- [11] A. Ridwan, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020.
- [12] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM," *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhmfMjtXw>
- [13] M. H. Asnawi, "Perbandingan Algoritma Naive Bayes, K-NN, dan SVM dalam Pengklasifikasian Sentimen Media Sosial", doi: 10.1234/pns.v10i.85.
- [14] T. Aura Azzahra *et al.*, "JURNAL MEDIA INFORMATIKA BUDIDARMA Perbandingan Efektivitas Naive Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," 2024, doi: 10.30865/mib.v8i1.7186.
- [15] D. Mualfah, R. Gunawan, and D. M. Suratno, "Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM," *J. FASILKOM*, vol. 13, no. 01, pp. 72–80, 2023.
- [16] M. Hudha, E. Supriyati, and T. Listyorini, "ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN#MATANAJWAMENANTITERAWA N DENGAN METODE NAIVE BAYES CLASSIFIER," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 1–6, 2022.

- [17] E. F. Saputra and M. R. Pribadi, “Analisis Sentimen Komentar Pada Kanal Youtube The Lazy Monday Menggunakan Algoritma Naive Bayes,” in *MDP Student Conference*, 2023, vol. 2, no. 1, pp. 17–23.
- [18] L. Lumbaa, “Implementasi Metode SVM dan Gradient Boost Dalam Klasifikasi Bahasa Daerah,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 908–915, 2022.
- [19] Endar Setyoroso, & Ismail. (2023). DESAIN UI/UX SISTEM INFORMASI MANAJEMEN MASYARAKAT TINGKAT RT BERBASIS WEB UNTUK MEMINIMALISIR MASALAH SOSIAL. *J-ENSISTEC (Journal of Engineering and Sustainable Technology)*, 10(01), 883–891.